A CONTRAST METRIC FOR 3-D VEHICLES IN NATURAL LIGHTING

G. Witus
Turing Associates, Inc.
1392 Honey Run Drive
Ann Arbor, MI 48103
USA
E-mail: witusg@umich.edu

G. Gerhart
U.S. Army Tank-automotive and Armaments Command
AMSTA-TR-R / MS 263
Warren, MI 48397
USA
E-mail: gerhartg@tacom.army.mil

1. SUMMARY

Ground vehicles in natural lighting tend to have significant and systematic variation in luminance over the presented area. This arises, in large part, from the vehicle surfaces having different orientations and shadowing relative to the source of illumination and the position of the observer. These systematic differences create the appearance of a structured 3-D object. 3-D appearance is an important factor in search, figure-ground segregation and object recognition.

This paper presents a contrast metric based on the 3-D structure of the vehicle, and an analysis of search performance for the Search_2 imagery. The analysis employs the traditional P-infinity-times-negative-exponential model of search time distribution. P-infinity and mean search time are modeled as functions of the target signature. The signature metric is one over the product of vehicle size and contrast. The value of the metric is measured by the ability to account for variance in observed search performance.

The 3-D structure contrast metric performs better than RSS contrast, and both perform dramatically better than the area-weighted average contrast. Target height performs better than either target area or square root of area. The signature metric accounts for over 80% of the variance in probability of detection and 75% of the variance in search time as measured in the TNO perception tests. When false alarm effects are discounted, the metric accounts for 89% of the variance in probability of detection and 95% of the variance in search time. The predictive power of the signature metric when it is calibrated to half the data and evaluated against the other half, is 90% of the explanatory power.

Keywords: Contrast ratio, 3-D perception, computational vision model, shape from shading, target acquisition, search

2. INTRODUCTION

Size and contrast have long been used to characterize the signature of simple targets in simple scenes for the purpose of analyzing search time and probability of detection. Size and contrast have been found to be good predictors of search and detection performance for stylized 2-dimensional targets, such as uniform disks and 4-bar patterns, against uniform backgrounds [Blackwell, 1943] [Ratches, et al., 1975].

Unfortunately, the standard area-weighted average contrast ratio has not proven to be a good predictor of search and target acquisition performance for complex targets in complex scenes. D'Agostino, et al., [1997] suggested a variety of possible modifications to the area-contrast metric to account for statistical luminance variation within the target and local surround. Peli [1996] concluded that the common measures of contrast are inadequate to explain detection performance for Gabor patches against uniform backgrounds, and suggested a computational contrast metric based on multi-scale band-pass filtering as an alternative.

Ground vehicles in natural lighting present non-uniform appearance when the surfaces of the vehicle are at different orientations with respect to the source of illumination and the observer (see fig. 1). The differences in shading between the adjacent surfaces reveal the 3-D structure. The appearance of common vehicles, from typical perspectives, under natural lighting is readily learned. This contributes to the perception of a 3-D object at a location, recognition of characteristic structural features, and classification as a potential vehicle.

Fig. 1. Example images of vehicles showing distinctive shading on different surfaces.

This paper presents the initial results of exploratory research to develop a contrast metric based on the 3-D vehicle structure, in natural lighting, relative to typical observer perspective. The objective of this research was to determine if a contrast metric could be defined based on the vehicle 3-D structure that would produce improved predictions of

Paper presented at the RTO SCI Workshop on "Search and Target Acquisition", held in Utrecht, The Netherlands, 21-23 June 1999, and published in RTO MP-45.

probability of detection and time to detect for military targets in natural backgrounds.

There is a substantial body of prior research suggesting that the perception of 3-D structure as a result of shape-from-shading is a significant factor in visual search and target acquisition. (Depth perception from visual parallax is insignificant at tactical ranges. For a stationary observer and a stationary target, shading and prior knowledge of vehicle appearance are the primary factors in 3-D shape perception.)

Marr [1982] coined the term "the 2½-D sketch" to refer to the perception of a 3-D structure from surface primitives. Sun and Perona [1996] showed that 3-D shading produced "pop-out" detection (i.e., response time independent of the number of distracters, indicative of pre-attentive parallel processing). They also showed that search became serial (time linear in the number of distracters) when the 3-D shading was removed even though the edge structure was retained. Tarr and Kersten [1998] concluded that the human visual system uses illumination angles to extract 3-D shape, and that illumination effects (including shadows) are modeled with respect to object shape, rather than simply encoded in terms of their effects in the image. Jonides and Gleitman [1972] and Mack and Rock [1998] both demonstrated that pre-attentive object recognition directs visual attention. Liu, Knill and Kersten [1995], and Liu and Kersten [1998] found that human efficiency exceeded 100 percent of an ideal 2-D observer for 3-D object classification. Moore and Cavanagh [1998] demonstrated that perception of 3-D shape is possible from limited surface shading information, given familiarity with the 3-D object. Ullman [1996] has shown that observers use 3-D operations to recognize familiar objects presented in novel orientations. 3-D surface matching is also an approach being pursued for automatic object recognition systems designed to work in clutter with partially occluded targets (e.g., [Johnson and Hebert, 1998]).

3. MODELING APPROACH

3.1 Contrast, Size and Signature Metrics

This exploratory investigation employed a simplistic, low-resolution approach. If 3-D shading is a significant factor in search and target discrimination, then the effects should be apparent even though coarse analytic techniques were used. If coarse analysis does not reveal an effect, then the effect, if present, is probably not strong enough to be worth addressing in search and target acquisition models.

The conceptual 3-D vehicle model was based on the 3 cardinal surface orientations of a rectangular solid (vertical front, vertical side, and horizontal top). While military vehicles are not rectangular solids, the 3-region geometric model can be adapted with a little work. The projected view of a vehicle was divided into the following three regions (see figure 2):

1. Front/rear. The near-vertical, negatively sloped or self-shadowed portion of the front (or rear depending on the presented aspect). For armored vehicles this includes the lower glacis, front track/tire, and turret-chassis gap. For trucks, this includes the front grill, front of the cab, and front of the tires.
2. Side. The near vertical (e.g., within 10 degrees), negatively sloped or self-shadowed portion of the side, including the sides of the tracks or tires.
3. Top. All horizontal and near-horizontal surfaces up to a slope of 80 degrees. This includes all the small miscellaneous objects and protrusions on the vehicle. It includes the the upper glacis, top deck, roof, rear deck, turret armor. It also includes the sloped rear roof of the HMMWV.

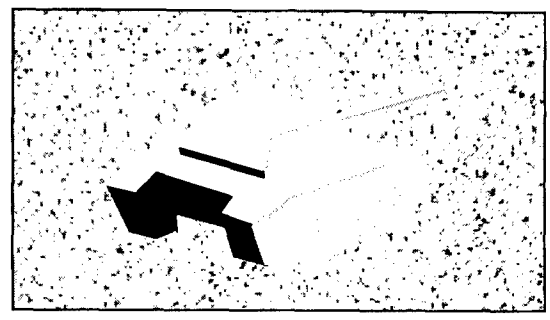

Fig. 2. Illustration of canonical front (rear), side, and top vehicle surfaces.

These canonical surfaces are meant to identify the major vehicle surfaces that typically have distinctive luminance resulting from different self-shadowing and angles relative to the observer and illumination. They address only the grossest level of 3-D structure. This level of resolution may be too coarse for modeling higher levels of target discrimination.

These regions also correspond to key structural features reported in field tests: darkly shadowed lower glacis, side profile, glint off roof or deck, lower grill, cab front, turret-chassis shadow. It is possible that the three surface orientations are significant because they correspond to important features for vehicle discrimination. It is also possible that the features are important because they reveal the 3-D structure.

The 44 Search_2 digital images [Toet, et al., 1998] were used in the demonstration analysis. All 44 images were used with no exceptions. The images were analyzed using Adobe Photoshop© to outline regions and compute gray-scale values. The local surround was taken to be a band surrounding the target with width equal to the target height.

The average gray-scale values for each region j, G_j, were converted to luminance values, L_j, via the calibration equation provided by Toet, et al., [1998]:

$$L_j = f(G_j) = 64.32 \left[(G_j - 18) / (G_j + 91.22) \right]^{2.3}$$

Since the calibration is a non-linear equation, a more accurate approach would have been to first convert pixel gray-scale to luminance, then compute the statistics.

The contrast for region j, C_j, is defined as the absolute value of the difference between the mean luminance of the region, L_j, and the mean luminance of the surround, L_{bkg}:

$$C_j = | L_j - L_{bkg} |$$

The vehicle contrast ratio metric, C_{veh}, is the area-weighted average of the contrasts of each of the three regions, divided by the luminance of the local background:

$$C_{veh} = \Sigma \, w_j \, Cj \, / \, L_{bkg}$$

where the weights, w_j, are the proportion of the presented vehicle area contributed by each region.

Two alternative contrast metrics were examined to provide a basis for relative comparison. These were the traditional area-weighted-average contrast [Ratches, et al., 1975] and the RSS contrast [D'Agostino, et al., 1997]. Both were computed by applying the non-linear gray-scale to luminance transform, f(), to statistics computed on the gray-scale images.

Signature metrics based on the area-weighted average contrast were uncorrelated with search performance (r^2 on the order of

0.3). Area weighted average contrast is not addressed in the remainder of this paper. This contrast metric was rejected.

The RSS contrast metric has been found to be an effective metric in other studies [D'Agostino, et al., 1997]. It is used as a reference for comparison with the 3-D structure contrast.

The RSS contrast ratio is the root-sum square of the target-background luminance difference and the target luminance standard deviation, divided by the mean background luminance:

$$RSS = [(\mu_{tgt} - \mu_{bkg})^2 + \sigma_{tgt}^{\ 2}]^{1/2} / \mu_{bkg}$$

For this comparison, the luminance standard deviation was estimated from the gray-scale mean and variance:

$$\sigma_{tgt} = [f([\mu_g^2 + \sigma_g^2]^{1/2})^2 - f(\mu_g)^2]^{1/2}$$

where f() is the gray-scale-to-luminance calibration equation.

The signature metric, S_{veh}, used in the analysis is simply one divided by the product of the vehicle size measure, V_{veh}, and vehicle contrast measure, C_{veh}:

$$S_{veh} = 1 / (V_{veh} C_{veh})$$

The size metric in this analysis was the target minimum dimension, generally the vertical extent or height. Vehicle height was the measure of size used in the early Night Vision Laboratory target acquisition modeling [Ratches, et al., 1976]. Target height (vertical extent) was reported in the Search_2 documentation.

Two alternative size metrics have been proposed as alternatives to target minimum dimension: the vehicle presented area, and the square root of the presented area [D'Agostino, et al., 1997]. These size metrics were examined, but their performance was inferior to target height. Only analysis results using height are presented.

3.2 Search Model

The analysis employed the traditional search performance model that expresses probability distribution of detection over time as the product of a limiting probability of detection (P_{inf}) and a negative exponential distribution:

$$P_d(t) = P_{inf} * (1 - e^{-(t-\varepsilon)/Td})$$

where e is the minimum reaction time (nominally 0.5 seconds) and T_d is the mean time to detect given that a detection occurs [Washburn, 1981] [Ratches et al., 1975].

In the perception test subjects were given 60 seconds in which to search and respond [Toet, et al., 1998]. Toet reports the mean search time plus reaction time. To obtain the parameters of the search model, the effects of the 60-second response window and reaction time must be discounted. Assuming the negative exponential distribution of search time, given that a detection occurs, mean search time, discounting windowing and reaction time, can be computed from the reported mean search time, T_s:

$$T_d = (T_s - \varepsilon) / (1 - e^{-(60-\varepsilon)/Ts})$$

Toet et al. [1998] also reports the number of detections, N_d, false alarms, N_f, and misses, N_m. Probability of detection within 60 seconds can be calculated from this data:

$$P_d(60) = N_d / (N_d + N_f + N_m)$$

In the test image set, only one image had $P_d(60)$ less than 0.4, three images had $P_d(60)$ less than 0.5, five images had $P_d(60)$ less than 0.6, and 24 images had $P_d(60)$ greater than 0.95. Figure 3 shows the relative number of detection, false alarm and time-out (miss) responses in the perception test.

P_{inf} is computed from T_d and $P_d(60)$

$$P_{inf} = P_d(60) / (1 - e^{-(60-\varepsilon)/Td})$$

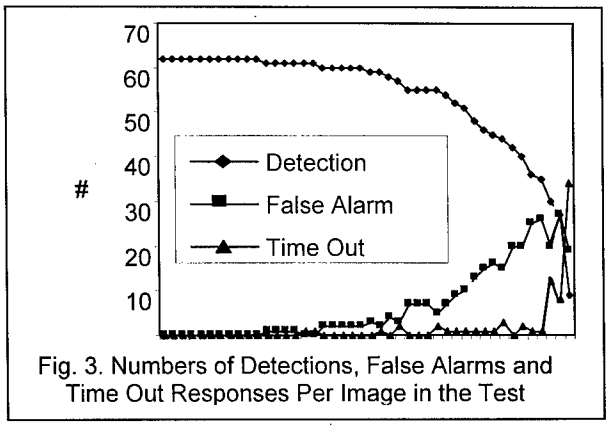

Fig. 3. Numbers of Detections, False Alarms and Time Out Responses Per Image in the Test

Given an unlimited search time, there are three possible outcomes: the observer can detect the target, false alarm, or conclude that there is no detectable target in the scene. Each of these is an absorbing state. As soon as the observer enters any one of these states, the search is over. Whenever a detection occurs, it is conditioned on having occurred before a false alarm and before the observer concludes that there is no target in the scene. In order for the conditional time to detect to have a negative exponential distribution, two criteria must be met:

(1) target detection, false alarms, and conclusion that no target is in the scene must be independent processes; and

(2) each of these processes must have a negative exponential distribution (albeit with different rates).

Under these conditions, the mean time to detect, conditioned on detection occurring first, is one over the sum of the individual rates of detection, R_d, false alarm, R_f, and concluding no detectable target is present, R_c:

$$T_d = 1 / (R_d + R_f + R_c)$$

P_{inf} is simply the ratio of the rate of true detection to the combined rates:

$$P_{inf} = R_d / (R_d + R_f + R_c)$$

These rates can be computed from the available data:

$$R_d = (1 / T_d) P_{inf}$$

$$R_f = (1 / T_d) N_f / (N_f + N_d)$$

$$R_c = (1 / T_d) - R_d - R_f$$

Cinlar [1975] and Washburn [1981] provide details of the mathematics of competing Markov processes.

The perception test in which the Search_2 data was collected used 35 mm slides with targets present in every scene. The subjects knew that each scene contained a vehicle. The subjects also knew that they had only 60 seconds in which to search the scene. Under these conditions, the subjects would presumably continue searching for the full 60 seconds. Since they knew a target was present, they would not conclude no detectable target was present within the first 60 seconds. This implies that R_c should be zero.

This hypothesis is supported by the data. The mean value of R_c, computed over the 44 images, is 0.0008, the maximum is 0.008, and the standard deviation is 0.0021. The expected time to conclude no detectable target is present ($1/ R_c$) is 21 minutes, and the standard deviation of the rate is 2.6 times the mean. R_c is not statistically significantly different from zero, and even if it was, it is so small as to be insignificant for this analysis. The remainder of the paper disregards R_c.

Figures 4 and 5 show the distribution of the rate of target detection, Rt, and the rate of false alarm, Rf. Note that the scales on the two graphs are an order of magnitude apart. Rt is greater than Rf for 43 of the 44 scenes.

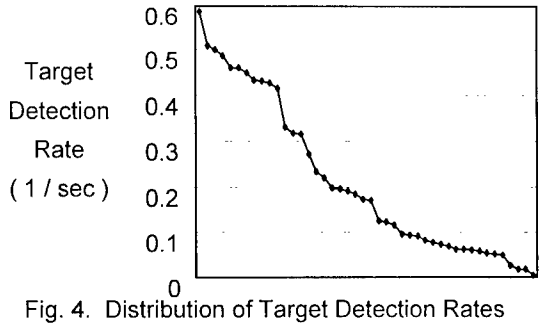

Fig. 4. Distribution of Target Detection Rates

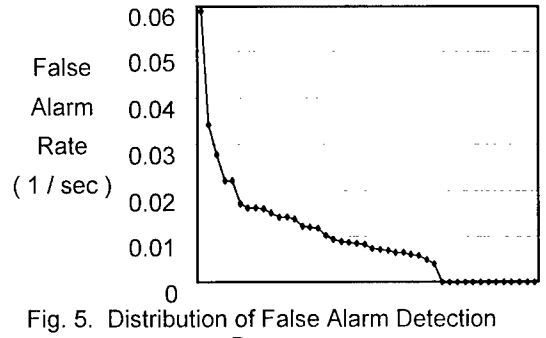

Fig. 5. Distribution of False Alarm Detection Rates

The mean time to detect a target, given that a detection occurs, discounting the effect of false alarms is

$$T_t = 1/R_d$$

When the effects of false alarms are discounted, P_{inf} has value one. The probability that a detection occurs within 60 seconds, discounting the effect of false alarms can be computed directly from the response data on the numbers of detections and missed targets

$$P_d(60 \mid \text{no false alarms}) = N_d / (N_d + N_m)$$

When the effects of false alarms are discounted, P_{inf} has value one. The probability that a detection occurs within 60 seconds, discounting the effect of false alarms can also be estimated from the computed from the unconstrained mean time to detect a target without false alarms, Tt:

$$P_d(60 \mid \text{no false alarms}) = (1 - e^{-(60-\varepsilon)Rd})$$

The root-mean-square (RMS) difference between these two estimates is 0.036, comparable to the sampling error in $P_d(60)$.

P_{inf} and T_d are modeled as simple linear functions of the signature metric. The model parameters (slope and intercept) are estimated from the data via linear regression. The related measures of search performance ($P_d(60)$, $P_d(60 \mid$ no false alarms), T_s and T_t) are modeled as functions of P_{inf} and T_d using the preceding search model equations.

4. ANALYSIS RESULTS

4.1 Gray-Scale Variance in the Vehicle Image

Partitioning the projection of the vehicle into the front, side and top regions accounted for 63 percent of the gray-scale variance over the entire target region. The area-weighted sum of the gray-scale variance within the three regions was 37 percent of the gray-scale variance over the entire target region. This indicates that these vehicle regions account for a significant proportion of the gray-scale variance in images of ground vehicles. Sources of residual variance include small features defining local surfaces with different orientations and self-shadowing, paint patterns, shadows from trees falling on the vehicle, and patches of foreground obscuration.

The RSS contrast metric includes all variance over the vehicle, regardless of structural significance or spatial scale of the variation. The RSS contrast and 3-D structure contrast have a strong statistical linear relationship ($r^2 = 0.90$).

4.2 Sampling Error in the Search Performance Data

Sampling errors are inherent to any test procedure with a finite number of subjects. If the identical experiment were repeated with different subjects, the results would differ due to sampling error and the stochastic nature of search and detection.

$P_d(60)$ is estimated as the proportion of observers correctly detecting the vehicle. Assuming observer responses are independent, the sampling error has a binomial distribution. For a given image, the one-sigma sampling error in $P_d(60)$ is given by the following equation:

$$\sigma_{Pd} = [P_d(60) * (1 - P_d(60)) / N]^{1/2}$$

where N is the number of subjects (N = 62).

Figure 6 shows a plot of sampling error in $P_d(60)$ versus observed $P_d(60)$ for the 44 Search_2 images. The RMS sampling error in $P_d(60)$ over the entire Search_2 image set is 0.0363. The standard deviation in measured $P_d(60)$ over the entire image set is equal to 0.187. Sampling error explains 3.8 percent of the variance in $P_d(60)$ over the image set.

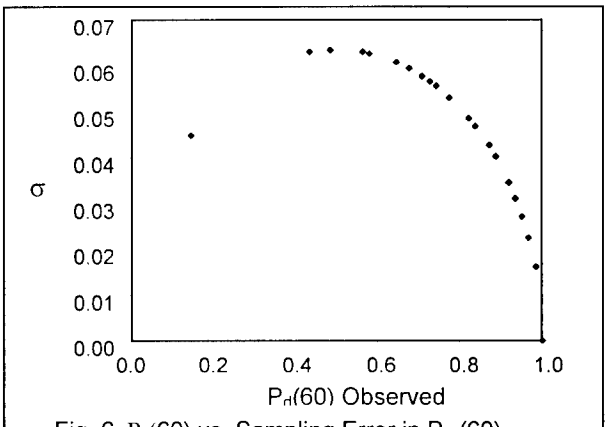

Fig. 6. $P_d(60)$ vs. Sampling Error in P_d (60).

The long-run probability of detection, P_{inf}, was not measured directly, but was computed from measured data. This makes the effects of sampling error difficult to compute. However the effects of sampling error can be approximated by assuming the random variables were measured. The sampling error in P_{inf} is 0.036, explaining 4.4% of variance.

The probability of detection in 60 seconds absent the effects of false alarms, $P_d(60 \mid$ no false alarms) is computed directly from the recorded data. The sampling error in $P_d(60 \mid$ no false alarms) is 0.025, explaining 3.8% of variance.

The mean search time reported, T_s, is a constant reaction time ($\varepsilon = 0.5$ sec) plus a random variable with a negative exponential distribution truncated at 60-ε seconds. For this analysis the standard deviation is approximated by the standard deviation of a negative exponential random variable with the same mean (i.e., no truncation). The standard deviation of a negative exponential random variable is equal to the mean. Since T_s is computed using response time data only for subjects who detect the vehicle, for any given image

the sampling error is equal to T_s divided by the square root of the number of subjects who correctly detected the vehicle:

$$\sigma_{Td} = (T_s - \alpha) / [\, N\, P_d(60)\,]^{1/2}$$

Figure 7 shows a plot of sampling error in T_s versus T_s for the 44 Search_2 images. The RMS sampling error in T_s over the entire Search_2 image set is 2.4 seconds. The standard deviation in measured T_s over the entire image set is equal to 7.58 sec. Sampling error explains 10.3 percent of the variance in T_s over the image set.

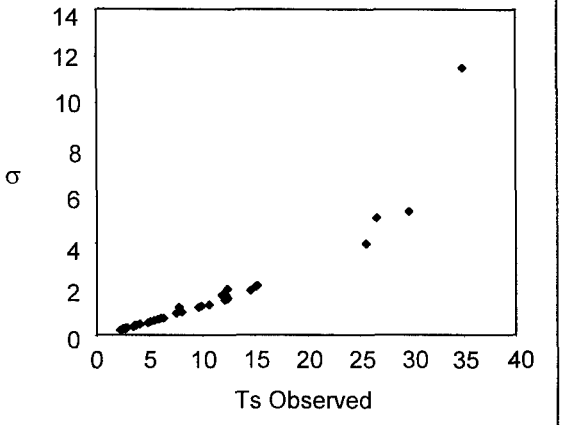

Fig. 7. T_s Obs vs. sampling error in T_s

The unconstrained mean time to detect, T_d, and the unconstrained mean time to detect in the absence of false alarms, T_t, were not measured directly, but were computed from measured variables. This makes the error due to sampling difficult to compute. The effects of sampling error can be approximated by assuming the random variables were measured. Both random variables have negative exponential distributions, so the 1-sigma sampling error is equal to the mean divided by the square root of the number of responding subjects. The estimated sampling error in T_d is 2.7 seconds, explaining 9.8% of variance. The estimated sampling error in T_t is 11.3 seconds, explaining 11.5% of variance.

4.3 Model Explanatory Power

The model has four free parameters that must be estimated from data: the slope and intercept of P_{inf} as a function of the signature metric, and the slope and intercept of T_d as a function of the signature metric.

The explanatory power of the model is measured by the percentage of variance in the observed search performance accounted for by the model. This is computed from the root-mean-square error between the model and observed data, and the variance in the observed data:

$$\%Var = 100\,(1 - RMS\_Error^2 / Observed\_Variance)$$

For a linear fit with parameters estimated via linear regression, the percentage of variance explained is equal to 100 times the Pearson correlation coefficient squared (r^2). Since the search model equations are non-linear, the percentage of variance accounted for is computed from the RMS error.

Figure 8 shows a scatterplot of the mean time to detect a target, given that the target is detected before a false alarm, but unconstrained by the 60 second time window of the experiment. The experimental value of T_d is computed from the measured search time. The model estimate of T_d is a linear function of the signature metric fit to the observed T_d.

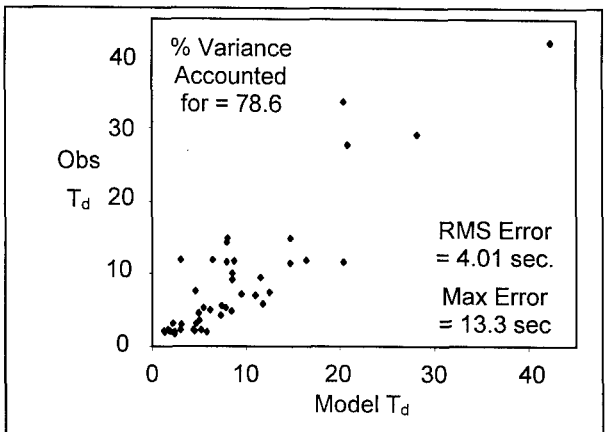

Fig. 8. Unconstrained Mean Time to Detect, T_d

Figure 9 shows a scatterplot of P_{inf} computed from the observed test data versus the linear function of the signature metric fit to the observed P_{inf} and truncated at one. Experimental values of Pinf are computed from T_d and raw response tallies.

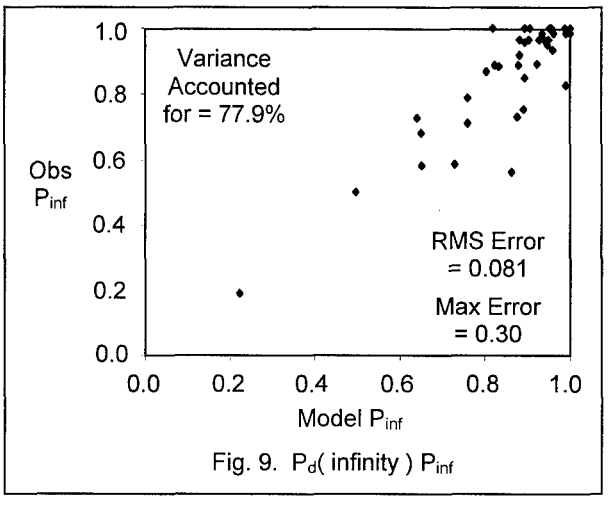

Fig. 9. P_d(infinity) P_{inf}

Figure 10 shows a scatterplot of the probability of target detection in 60 seconds computed directly from the tallies of observer detections, false alarms and misses, versus the model $P_d(60)$ computed from P_{inf} and T_d. The results are very similar to the Pinf results because, in most cases, the mean search time was much less than the 60 second response window.

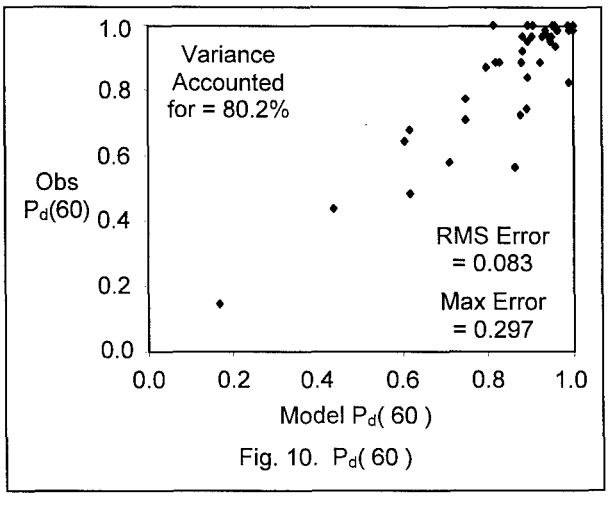

Fig. 10. P_d(60)

Figure 11 shows a scatterplot of the mean search time measured in the experiment, versus the mean search time calculated by the model accounting for the effects of competing false alarms and the 60 second response window. These results resemble the results for unconstrained search time because, in most cases, the mean search time was much less than the 60 second response window.

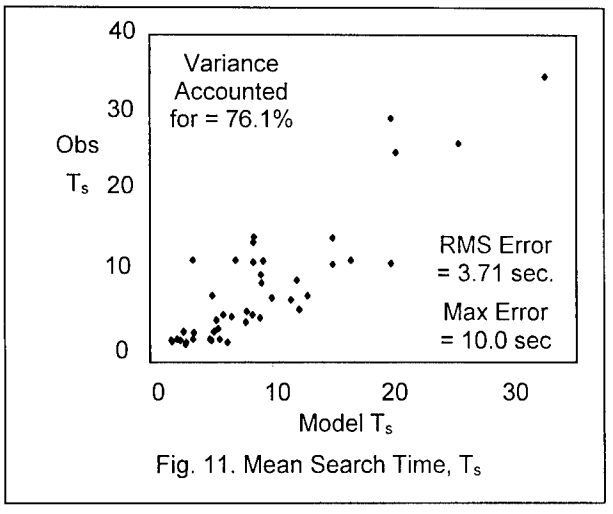

Fig. 11. Mean Search Time, T_s

Figure 12 shows a scatterplot of the mean time to detect a target without the requirement that the target detection occurs before a the first false alarm, T_t. It is the inverse of the rate of target detection. It is computed as T_d divided by P_{inf}. The experimental and model values of T_t are computed from the experimental and model values of values of T_d and P_{inf}. When the RSS contrast is used instead of the 3-D structure contrast, the percent of variance accounted for drops from 95% to 89%.

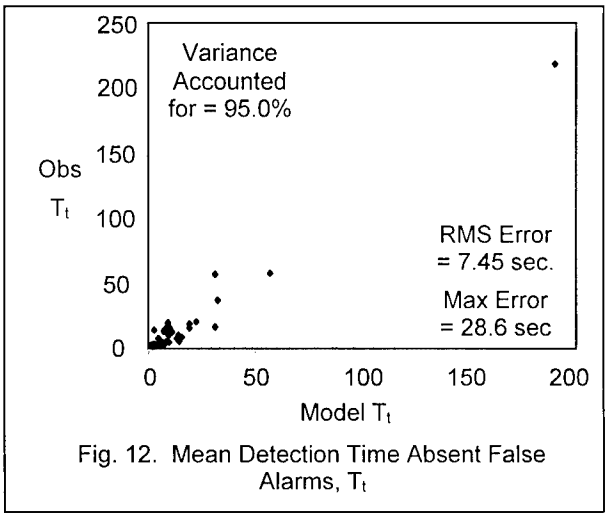

Fig. 12. Mean Detection Time Absent False Alarms, T_t

Many of the data points in figure 12 are clustered near the origin. The correspondence for low response time cases is seen more clearly when the logarithm of T_t is plotted (see figure 13). The logarithm operation is a non-linear transformation, so the percent of variance accounted for is different.

Interestingly, the percent of variance accounted for by Ln(Model T_t) is equal to the percent of variance accounted for by linear regression of the signature metric directly on Ln(Observed T_t). When the RSS contrast is used instead of the 3-D structure contrast, the percent of variance accounted for drops from 76% to 50%.

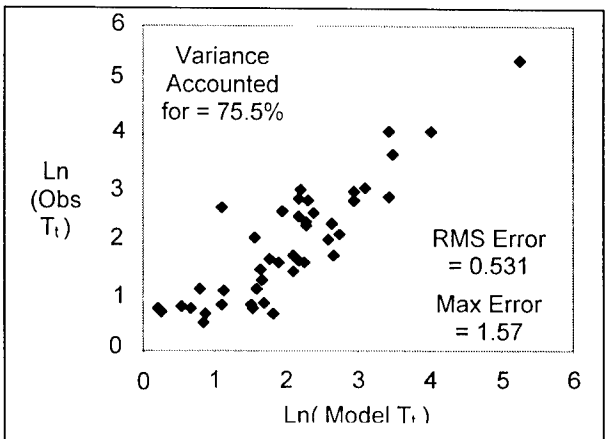

Fig. 13. Ln (Mean Detection Time Absent False Alarms), Ln(T_t)

Figure 14 shows a scatterplot of the probability of target detection in 60 seconds, without competition from false targets, i.e., excluding false alarms. The experimental value is computed from the tallies of detections and misses. The model value is computed from T_t.

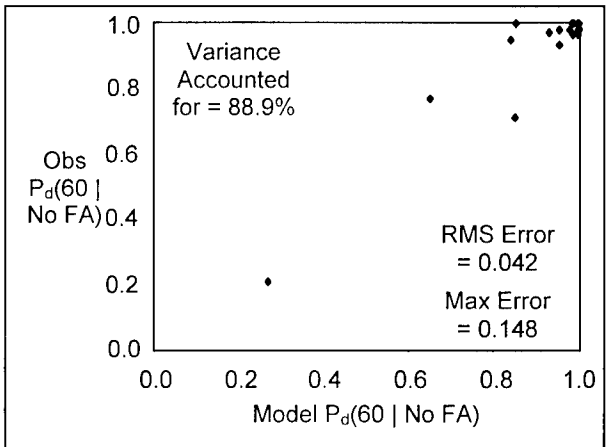

Fig. 14. P_d(60 | No False Alarms)

Tables 1 and 2 summarize the results of the comparison of the model to the data, and compare results obtained using the 3-D structure contrast metric with those obtained using the RSS contrast metric. Table 1 presents the percent of variance explained by the model. Table 2 shows the magnitude of the maximum model error.

Search Performance Measure	% Var 3-D	% Var RSS
Unconstrained Time to Detect, T_d	78.6	77.8
P_{inf}	77.9	76.0
Search Time, Ts	76.1	75.2
P_d(60)	80.2	78.6
Detection Time Sans F.A., T_t	95.0	88.5
P_d(60 | No False Alarms)	88.9	86.5

Table 1. Model Explanatory Power

Search Performance Measure	Max Error 3-D	Max Error RSS
Unconstrained Time to Detect, T_d	13.3	11.9
P_{inf}	0.30	0.31
Search Time, Ts	10.0	9.0
P_d(60)	0.30	0.31
Detection Time Sans F.A., T_t	28.6	60.2
P_d(60 | No False Alarms)	0.15	0.19

Table 2. Maximum Model Error

The results have several significant implications:

1. Signature metrics based on both the 3-D structure contrast metric and on the RSS contrast metric account for large proportions of the variance in search performance for this data set.
2. The 3-D structure contrast metric accounts for one to two percentage points more variance than the RSS contrast metric, except for the mean time to detect, absent false alarms where there is a 6.5 percentage points difference. This indicates that the 3-D structure contrast is a better measure observer response to the target signature. When the effects of false alarms are included, the additional variance due to this non-target source obscures the difference between the two contrast metrics.
3. The percentage of variance predicted by both metrics is significantly higher when the effects of false alarms are discounted. This is not surprising since the signature metrics do not measure potential false targets. The difference is greater for 3-D structure contrast than for the RSS contrast, further supporting the claim that 3-D structure contrast is a better measure of the effects of the target signature.
4. The difference in the percent of variance predicted with and without the effects of false alarms indicates the magnitude of the contribution of false targets to search performance variance. By this measure, false targets account for over 15% of the variance in the mean time to detect a target, and almost 9% of the variance in the probability of target detection within 60 seconds.
5. The maximum error in P_d(60 | no false alarms) is significantly lower that the maximum probability error when the effects of false alarms are not excluded.
6. The magnitude of the maximum detection time sans false alarms is large. However this error occurs at the one hard-to-detect image, for which T_t was 218 seconds. The error, as a percentage of the time for that data point, is 13% for the 3-D structure contrast metric and 28% for the RSS metric.

Several excursions were conducted to assess alternative vehicle size metrics. When the signature metric was calculated using the square root of the presented area instead of the target height, the percent of variance predicted was approximately 12 percentage points lower for P_{inf} and 3 percentage points lower for T_d. When the presented area was used, the results were 15 percentage points lower for P_{inf}, and 6 percentage points lower for T_d.

4.4 Signature Metric Measurement Error

Measurement error occurs because the original images were blurred. The boundaries of the vehicles and regions in the vehicles were not sharply delineated. This affected both the measurement of target vertical extend and luminance. Not only was the location of the boundary uncertain, but pixels near the boundary contained a mix of target and background luminance, or a mix of the luminance between two regions.

Two separate estimates of the 3-D structure contrast ratio were made. Toet et al., [1998] provided one measurement of target height. A second independent measurement was made in this study. These measurements provided two pair of independent measures of the signature metric. Each independent pair of estimates produced one estimate of the measurement error in the signature metric.

The one-sigma measurement error in the signature metric over the Search_2 image set is 0.016. Since the model is linear, the measurement error in the predictions of P_{inf} and T_d are 0.016 times the magnitude of the slope (-2.313 and 92.11 respectively). This analysis yields a one-sigma measurement error in the predictions of 0.036 for P_{inf} and 1.473 for T_d respectively. The measurement errors in the predictions of P_{inf} and T_d are less than the sampling errors in the perception test estimates of P_{inf} and T_d (0.036 and 2.7 seconds respectively).

In combination, the variance due to sampling error and signature metric measurement error together are 9.1 percent of the P_{inf} variance predicted by the model, and 18.5 percent of the variance in T_d predicted by the model. The predictive power of the signature metric cannot be the result of spurious sampling and measurement errors.

The signature metric is one over the product of the vehicle 3-D structure contrast and the vehicle height. Two measurements of height and contrast were made, to obtain two pairs of independent measurements of the signature metric. The two correlations between the two pair of signature metric measurements were 0.986 and 0.979. The sample standard deviation for a pair of independent measurements is simply the difference between them. The error estimate for two pairs of independent measurements is the RMS of the two estimates of the sample standard deviation.

Figure 15 presents a plot of the signature metric measurement error sample standard deviation versus the signature metric value. The correlation is 0.91, suggest a strong linear relationship with slope equal to 0.15. As expected, the measurement error is larger for small, low-contrast vehicles than for large, high-contrast vehicles.

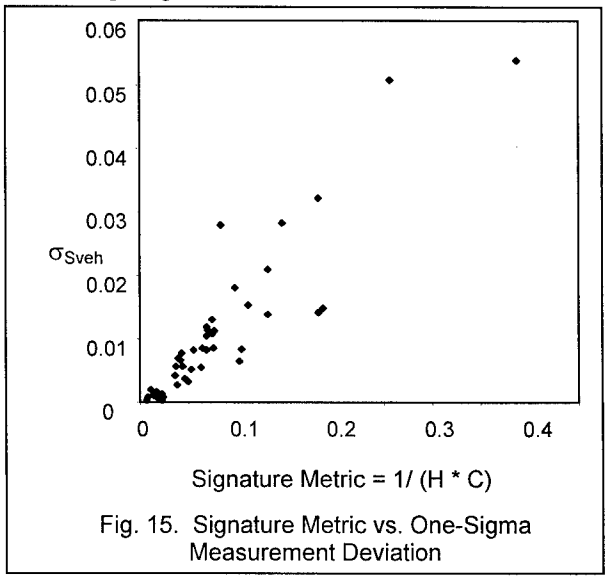

Fig. 15. Signature Metric vs. One-Sigma Measurement Deviation

4.5 Accounting for Residual Variance

The model accounts for 75 to 80 percent of the variance in the experimental data when the effects of false alarms are included, and 90 to 95% of the variance when they are discounted. This suggests that for the TNO Search_2 data and test, false alarms account for 10 to 15 percent of the variance in probability of detection and search time, respectively.

Sampling error accounts for approximately 4 percent of the variance in probability of detection and 10 percent of the variance in search time. Together the target signature, false alarms, and sampling error are sufficient to account for all of the variance in the experimental data. (It is not possible simply to sum the percentage of variance explained by sampling error with the percentage of variance explained by the signature metric because of spurious correlation when the model parameters were estimated from the data).

The signature metric was calculated by applying the non-linear gray-scale-to-luminance transform to the mean and RMS gray-scale values. The correct method is to apply the gray-to-luminance transform to the image, then compute statistics. This approximation may account for some of the residual variance.

The contrast metric did not include any measure of color contrast or texture differences. The metric did not address chromatic, luminance or contrast adaptation, or spatial filtering. The metric did not address the effect of the position of the target in the scene, or its position relative to other features that might attract of inhibit attention to the target location. These factors may contribute to the model error, but the effect is likely to be small because the unexplained error is small.

There is no term that can be added to the signature metric to account for the residual variance. The prediction errors in P_{inf} and T_d are only weakly correlated ($r^2 = 0.25$). This suggests that once target signature effects are discounted, probability of detection and search time are sensitive to different processes and/or have non-linear relationships with image characteristics.

4.6 Individual Effects of Size and Contrast

One over the 3-D structure contrast metric was modestly correlated with perception test data ($r^2 = 0.7$ for T_d and 0.6 for P_{inf}). The percentages of variance explained for T_t and P_d(60 | no false alarms) were 51% and 54%. The RSS contrast metric had comparable correlation to T_d and P_{inf}, but accounted for 10 percentage points less of the variance in T_t and and P_d(60 | no false alarms). The area weighted average contrast ratio had essentially no correlation with T_d or P_{inf}.

Target height, area and square root of area were only weakly correlated with T_d and P_{inf} (r^2 approximately equal to 0.4). Height had some correlation with T_t and P_d(60 | no false alarms) with r^2 on the order of 0.2. For area and square root of area, accounted for less than 10 percent of variance in T_t and and P_d(60 | no false alarms).

Height was less correlated with the 3-D structure contrast metric than it was with the RSS contrast metric ($r^2 = 0.3$ for the 3-D structure contrast metric, versus 0.4 for the RSS contrast metric).

These data indicate that height and 3-D structure contrast were largely independent dimensions, which individually were moderately correlated with search performance. Not surprisingly, their product was well-correlated with search performance. The same statements are true to a lesser extent for the RSS contrast metric.

4.7 Spurious Correlation and Predictive Power

When the same data are used to calibrate and evaluate the model, the percentage of variance accounted for is an accepted measure of explanatory (descriptive) power, but it is not truly a measure of the model's predictive power. In order to assess the model's predictive power, the model must be calibrated to one data set, then the prediction error evaluated for a separate, sequestered data set. This minimizes the effects of spurious correlation.

The Bootstrap statistical technique [Davison, 1997] was used to evaluate the predictive power of the signature metric. The Bootstrap technique involves repeated random partitioning of the data into two disjoint sets: the calibration data set containing, and the validation data set. The model parameters are estimated from the calibration set, then the RMS prediction error is calculated from the sequestered validation data set. Each partition produces an estimate of the variance predicted by the model.

In this particular application of the Bootstrap technique, the calibration and validation data sets each contained half of the data points. Twenty-two calibration data points were used in the linear regression to estimate the model parameters, and 22 data points in the validation set were used to measure the RMS error and the percent of variance in the validation data set predicted by the model. Two-hundred-fifty-two (252) random partitions were generated to compute the Bootstrap statistics.

The Bootstrap analysis was applied to investigate the ability of the signature metric to predict the logarithm of the mean time to detect a target in the absence of false alarms, T_t. T_t was chosen as the dependent variable because it had the clearest causal relationship to the signature metric. The logarithm of Tt was used because of the uneven distributions of observed T_t and for the signature metric (see figures 12 and 13).

Figure 15 shows the distribution of slope and intercept from the 252 partitions. The slope had median value 14.6, equal to the slope when all 44 points are used in the regression (theBootstrap mean and variance are 15.0 and 1.6 respectively). The intercept has a median value of 0.89 compared to 0.90 when all 44 points are used in the regression (the Bootstrap mean and variance are 0.88 and 0.14 respectively).

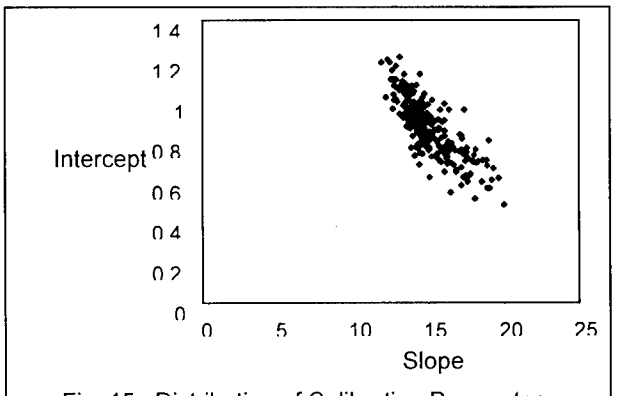

Fig. 15. Distribution of Calibration Parameter Values from Bootstrap Replications

Figure 16 shows the distribution of the percentage of variance in the calibration data sets versus the percentage of variance predicted in the validation data sets. The median percent of variance predicted in the validation data sets is 72% (the mean and variance are 70% and 10 percentage points, respectively). The median percent of variance explained in the calibration data sets is 78%, compared to 76% when all 44 points are used

in the regression (the Bootstrap mean and variance are 76% and 9 percentage points, respectively). On average (median and mean) the proportion of variance predicted in the validation data set is 92% of the proportion of variance accounted for in the calibration data set. This difference is due to spurious correlation, and indicates the difference between the explanatory and predictive power of the signature metric.

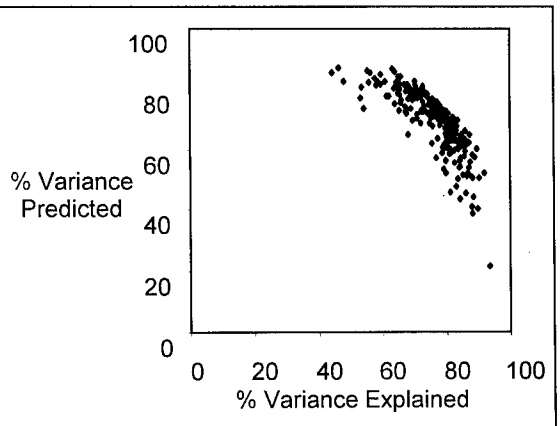

Fig. 16. Distribution of Percentage of Variance Explained and Predicted by the Signature metric

5. FINDINGS AND OBSERVATIONS

The traditional model of the distribution of search time was a useful model to analyze the experimental data.

With appropriate choice of definition of size and contrast, the simple signature metric equal to one over the product of size and contrast is a good fit to the observed data. It explains 75 to 80 percent of the variance in the test data, and 90 to 95 percent when the effects of false alarms are discounted.

The organization of the vehicles into three regions based on their orientation relative to the illumination and observer accounts for a significant portion of the gray-scale variance. Not surprisingly, the 3-D structure contrast metric and the RSS contrast metric are highly correlated and produce comparable results.

Nonetheless, the 3-D structure contrast metric is consistently superior to the RSS contrast metric, especially when the effects of false alarms are discounted. Variance due to false alarms obscures the difference in performance for the two contrast metrics. Both contrast metrics are far superior to the area weighted average contrast (which is no good at all).

Vehicle height is a better measure of target size, for use in product with a contrast metric, than either vehicle area or square root of vehicle area.

Responses to false targets, i.e., false alarms, account for a 10 to 15 percent of the search performance variance. Modeling the rate of false alarms as a function of the image properties has potential to improve search modeling.

There are a number of low-level and high-level visual phenomena not represented in this simple signature metric. Low-level factors include color contrast, chromatic and luminance adaptation, spatial filtering and contrast adaptation. Mid-level image processes include pre-whitening, edge detection and texture segregation. Beyond the vehicle structure, high-level (top-down) image properties include the location of the vehicles relative to terrain features that might attract attention or direct attention away, and position of the vehicle in the image.

These factors could account for the unexplained variance. However, they were not major contributors to search performance variance in the Search_2 image set. These factors could be more significant in other image sets containing greater variation on these dimensions.

The Search_2 vehicles do not present significant perceptible camouflage. Camouflage adds variance to the image. The RSS contrast metric will yield higher values for vehicles with camouflage than for vehicles without camouflage (assuming the same mean luminance), and thus will predict higher P_{inf} and lower T_d for camouflaged vehicles than for comparable non-camouflaged vehicles. The 3-D structure contrast metric is camouflage-neutral since it is based only on the mean luminance of different target regions and does not incorporate any higher-order statistics.

The Search_2 vehicles do, in some cases, have perceptible structures within the front, side and top regions. This increases P_{inf} and decreased T_d. These structures add variance to the image, which increases the value of the RSS contrast metric, which leads to higher predicted P_{inf} and lower predicted T_d. The 3-D structure contrast metric is neutral with respect to structures within the three regions.

Neither the 3-D structure contrast metric nor the RSS contrast metric are able to distinguish modulation due to internal structure from modulation due to camouflage or foreground obscuration (e.g., brush or nets). More sophisticated signature analysis is needed to make this distinction.

6. CONCLUSIONS

The 3-D structure of the vehicle is a promising basis for signature analysis. Basic research suggests that shape from shading and 3-D appearance are pop-out cues, focus visual attention, facilitate figure-ground segregation. This analysis provides evidence that 3-D structure is an important factor in search and target acquisition in natural settings.

The 3-D structure contrast metric was useful in analyzing the Search_2 image set. The simple approach explored in this paper may not be robust enough for a wide variety of image sets. Future research should explore extending the 3-D structure analysis approach and using it in combination with a computational model of front-end visual processing.

The traditional model of the distribution of time to detect a target was a useful framework with which to analyze search performance. The search model was extended to express P-infinity as a function of the rates of detection, false alarm and deciding that no detectable target is present. This extension enable the analysis to quantify the effects of false alarms on variance in search performance.

The signature metric had very strong explanatory power for this data set, especially when the effects of responses to false targets were discounted. Limited Bootstrap analysis suggests that the predictive power of the model is 92 percent of the explanatory power.

False alarms were a significant factor contributing to variance in search performance. Further research is needed to demonstrate effective models to predict the rate of false alarm from image properties and top-down knowledge.

The specific quantitative results of this analysis, especially the calibration of P_{inf} and T_d as linear functions of the signature metric, are unlikely to transfer to other perception tests and image sets. Observer response depends on the test-specific factors such as the proportion of images with no target, the relative penalty of false alarms versus missed detections, the response time window, search area, etc. Image sets that

contain different distributions of target signatures, false targets, scene complexity (terrain features), etc. will lead to different quantitative results.

7. ACKNOWLEDGEMENTS

This research was funded by the US Army Tank-Automotive Command Research Development and Engineering Center (TARDEC) under contract DAAE07-97-C-X101. The views and opinions expressed in this paper are those of the author and do not reflect the policy or position of any agency of the United States Government.

8. REFERENCES

1. Blackwell, H. R., "Contrast thresholds of the human eye," *J. Opt. Soc.* 36: 624-43, 1943.
2. Cinlar, E., *Introduction to Stochastic Processes*, Prentice Hall, 1975.
3. D'Augustino, J., W. Lawson and D. Wilson, Concepts for search and detection model improvements, *Proceedings of the SPIE* 3063: 14-22, 1997.
4. Davison, A. C., and D. V.Hinkley, *Bootstrap Methods and Their Application,* New York: Cambridge University, 1997.
5. Johnson, A. E. and M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image and Vision Computing* 16.9-10: 635-51, 1998.
6. Jonides, J. and H. Gleitman, "A conceptual category effect in visual search: O as a letter or digit," *Perception and Psychophysics* 12: 457-60, 1972.
7. Liu, Z. and D. Kersten, "2D observers for human 3D object recognition?" *Vision Research* 38.15-16: 2507-19, 1998.
8. Liu, Z., D. C. Knill, and D. Kersten, "Object classification for human and ideal observers," *Vision Research* 35.4: 549-68, 1995.
9. Mack, A. and I. Rock, *Inattentional Blindness*, Cambridge: MIT Press, 1998.
10. Marr, D., *Vision*, New York: W. H. Freeman & Co., 1982.
11. Moore, C. and P. Cavanagh, "Recovery of 3D volume from 2-tone images of novel objects," *Cognition* 67.1-2: 45-71, 1998.
12. Peli, E., "In search of a contrast metric: matching the perceived contrast of Gabor pathes at different phases and bandwidths," *Vision Research* 37.23: 3217-24, 1997.
13. Ratches, J. A. et al., "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems," *R&D Technical Report ECOM-7043,* April, 1975.
14. Sun, J. Y. and P. Perona, "Preattentive perception of elementary three-dimensional shapes," *Vision Research* 36.16: 2515-29, 1996.
15. Tarr, M. J. and D. Kersten, "Why the visual recognition system might encode the effects of illumination," *Vision Research* 38.15-16: 2259-75, 1998.
16. Toet, A. et al., "A high-resolution image data set for testing search and detection models," *TNO-Report TM-98-A020*, Soesterberg, The Netherlands: TNO Human Factors Research Institute, 1998.
17. Ullman, S., *High-Level Vision: Object recognition and Visual Cognition,* Cambridge: MIT, 1996.
18. Washburn, A. R., *Search and Detection,* Military Operations Research Society of America, 1981.